DOCUMENT RESUME

ED 144 406 FL 008 912

AUTHOR Callaway, Donn R.
TITLE Accent and the Evaluation of ESL Oral Proficiency. Occasional Papers on Linguistics, No. 1.
INSTITUTION Southern Illinois Univ., Carbondale. Dept. of Linguistics.
PUB DATE Apr 77
NOTE 16p.; Proceedings of the International Conference on Frontiers in Language Proficiency and Dominance Testing (1st, Southern Illinois University, Carbondale, Illinois, April 21-23, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Communicative Competence (Languages); *English (Second Language); Evaluation Methods; *Language Proficiency; Linguistic Competence; *Linguistic Performance; Oral English; *Oral Reading; *Rating Scales; *Reliability; Second Language Learning; Speech Communication
IDENTIFIERS Interrater Reliability; Rater Reliability

ABSTRACT

A study was carried out to ascertain whether language instructors or naive judges were more reliable in judging oral proficiency. Fifteen students were chosen from the ESL (English as a second language) center at Southern Illinois University to record a tape while reading passages in English. The tape and a questionnaire were administered to 70 raters. Half of them were students who had neither linguistic nor teaching experience (the naive group) while the other half were instructors or teaching assistants in ESL (the experienced group). Each sample was rated by the judges on each of six scales: the first four scales consisted of four pairs of bipolar adjectival descriptors, an overall proficiency scale, and a multiple choice questionnaire about the language background of the speaker. There was a very substantial agreement among the raters, regardless of whether they were naive or experienced. Both groups were very reliable on the whole in judging proficiency level, although the experienced were somewhat more reliable. The experienced raters were much better at identifying correctly the source language backgrounds than the naive group. (CFM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James E Redden, Editor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED144406

FL008912

Accent and the Evaluation of ESL Oral Proficiency

Donn R. Callaway

Southern Illinois University

The evaluation of oral proficiency in a second language has always been a formidable problem for the second language teacher. The teacher must accurately evaluate the learner's ability "to receive or transmit information in the test language for some pragmatically useful purpose" (Clark, 1975, 10).[1] However, in order to evaluate the student's oral proficiency the teacher must rely on highly subjective judgements of the student's output.

Studies of evaluative reactions to speech samples by Lambert (1960), Labov (1966), and others have supported the notion that a linguistically-naive listener can and does make critical evaluations of the speaker's personality, social class, or ethnicity. These studies have used either variations between dialects or between languages as the independent variable. While the dependent variables have consisted of a general personality factor, a social class rating, individual personality traits, or separate speech characteristics. Since, in these studies, speech characteristics tended to be ignored, Williams called for research:

> to link whatever language and speech features serve as salient cues in this judgmental process with whatever kinds of evaluation or stereotypes are of interest to us in the behavior of listeners (1970, 473).

Williams showed that social class and judgement ratings can be predicted from the presence, absence, or strength of certain language features, among them silent pausing and verb constructions (1970, 477). However, since the study was limited to the effect of Black dialect on Black and White elementary school teachers and did not include other dialects and other evaluators, further research is needed.

As a result of the concern of sociolinguists that accented speech can cause alienation and discrimination in educational and occupational opportunities (Ortego, 1970: Ryan, 1973), most studies have dealt only with the language varieties of various minority groups: French Canadians (Lambert, Hodgson, Gardner, and Fillenbaum, 1960; Anisfeld and Lambert, 1964; Webster and Kramer, 1968), Black Americans (Harms, 1963; Shuy, Baratz, and Wolfram, 1969; Tucker and Lambert, 1969; Williams, 1970; Williams, Whitehead, and Miller, 1971), Mexican Americans (Ortego, 1970; Williams, Whitehead, and Miller, 1971; Ryan, 1973), and British regionals (Strongman and Woosley, 1967; Giles, 1972). Richards points out that any:

> Deviancy from grammatical or phonological norms of a speech community elicits evaluational reactions that may classify

a person unfavorably (1971, 21).

If this divergence from the standard language has an effect on these minorities, then it could have an even more pronounced effect on the acceptance of the second language learner.

In oral evaluation, a general assumption is that any native speaker can assess the proficiency of a non-native. In fact, a generally used measure, the American Language Institute Oral Rating From, describes a speaker of minimum proficiency as having "pronunciation...virtually unintelligible to '*the man in the street* [my italics]' "(1962). However, usually, it is the second language teacher, rather than 'the man in the street' who makes the evaluation. Is it safe to assume that there is no significant difference between a trained rater and an untrained one? Cartier does not think it is. He says that judgements of proficiency

> are made by the wrong people, they are made by sophisticated language instructors who have become quite skilled at understanding heavily dialectal English rather than the student's eventual instructors, classmates, and job supervisors (1968, 21).

He seems to be implying that naive judges might be better. According to Jakobovits (1970), however, naive judges are apt to attribute too much importance to "accent, pronunciation, and fluency" and too little to the weightier matters of grammar and vocabulary (85).

A number of research studies have examined the effect of accent in bidialectal and bilingual speech. These projects have dealt with the reliability of judges, their ability to specify speech characteristics, or their ability to judge the degree of accentedness and the proficiency of the speaker.

In 1973, Gorosch compared oral EFL proficiency evaluations in Sweden by teachers, who were non-native speakers of English, and non-teachers, who were native speakers of English. His evaluators rated six Swedish EFL learners by noting each mistake in pronunciation and then they assessed overall intelligibility on a five-point scale. His data indicated that both groups tended to separate pronunciation from intelligibility and that the evaluations of the non-teachers were unpredictable. He concluded that there are "considerable differences between assessments produced by teachers and those produced by non-teachers" (151).

Contrary to Gorosch's findings, a study by Brennan, Ryan, and Dawson (1970) demonstrated that native speakers could give reliable judgements. Seventy-two naive listeners judged the degree of accentedness in 8 samples of Spanish-English. In addition to the high reliability, the results indicated that the judges agreed on what degree of accentedness constituted a proficiency level. Further, it showed that the subjects during

an informal question period were unable to articulate which features of language were important to their judgements.

Through a 1970 experiment, Giles attempted to arrange three supposed speech characteristics hierarchially: the aesthetic, the communicative, and the status. These three characteristics were defined as the pleasantness of the voice, the intelligibility of the accent, and the prestige value of the accent. His subjects, adolescents from S. Wales and S.W. England, listened to and rated 13 British regional and foreign accents on the three characteristics mentioned above. His data showed that although the subjects were able to identify the individual accents, the three characteristics were apparently indistinguishable. He concluded that these characteristics were at best "three variants of one evaluative dimension" (219).

Further research conducted by Galvan, Pierce, and Underwood (1975) examined the speech of Mexican-American bilinguals in terms of 10 personality traits and 10 speech characteristics. Five recorded samples were evaluated by 92 American undergraduates. The analysis showed that the raters generally evaluated speakers more positively than negatively, that the evaluations became more negative as accentedness increased, and that few differences could be predicted on the basis of the listeners' academic backgrounds. In regards to the scales of speech characteristics (relaxed, appropriate, natural, standard, graceful, careful, understandable, good English, active, and smooth), they speculated that not only could the characteristics be reduced to a few factors but from these factors the points on an accent continuum could be established (15).

The previously mentioned studies all had one thing in common. They dealt with the speech of fairly established groups, whether these groups were bidialectal or bilingual. A study by Palmer (1973), however, dealt with a transient group of learners. Palmer produced a preliminary report of the subjective evaluations of ESL learners by naive listeners. Eighteen students from Georgetown Univesity listened to 36 speakers from four different language backgrounds (Lingala, Arabic, Spanish, and Vietnamese). Each judge evaluated each speaker on a five-point scale across three tasks: reading, retelling, and narration. The judges reliably rated the speakers across language backgrounds and across tasks. However, the raters were not very good in identifying the source language background of the speaker. It would seem from his results that particular foreign accents, i.e. Spanish, may not be as distinctive as popularly believed.

This study attempts to address a number of remaining questions. Here, as in Palmer's study, the focus is on learners of English as a second language; but unlike Palmer's study, this one compares interrater reliability among naive raters with interrater reliability among experienced ESL teachers. Further, evaluations by both groups are evaluated against the independent placement of the speakers in

one of five proficiency levels, via a separate testing procedure. Whereas the earlier studies were concerned with the effect of a speaker's accent on the listener's judgement without knowing what characteristics of speech contributed to that evaluation, this experiment will attempt to separate accentedness into distinguishable dimensions.

In particular, the following questions are addressed:

(1) How much agreement is there on the evaluation of samples of non-native speech among judges?
(2) Is there any difference between naive native raters and experienced ESL teachers in the evaluations of foreign speech?
(3) Will evaluations of non-native speech samples correlate with the independent placement of the speakers in an intensive English program, that is, do the ratings have concurrent validity?
(4) Are there distinguishable dimensions for the ratings of accentedness in this study?
(5) How accurately can the judges identify the source language background of each speaker?

METHOD

Speech Samples. The experimenter asked instructors from each of the five proficiency levels at the Center for English as a Second Language (CESL), Southern Illinois University, to recommend students of average speaking ability from Arabic, Persian, and Spanish speaking language backgrounds who would be willing to participate in a short recording session. Two native speakers who were graduate students in the Department of Linguistics, Southern Illinois University, were also taped. Originally 25 ESL students, along with the two American students, were recorded in a laboratory setting reading one of twelve 100-word passages in English. (The paragraphs are given in Appendix A.) Each speaker was allowed to practice the passage twice before he was recorded. Eleven tape samples were eliminated due to excessive noise or to speaker duplication. Three non-natives from each of the five levels of CESL (15 ESL students) were finally selected along with one of the American students.

Questionnaire: Scales of Accentedness and Overall Proficiency. The scales were constructed in a semantic differential form similar to the scales used by Lambert, Galvan, Palmer, and their colleagues. The first four scales consisted of four pairs of bipolar adjectival descriptors, "not very intelligible" to "intelligible," "unpleasant" to "pleasant," "unacceptable" to "acceptable," and "non-native" to "native." In addition, there was an overall proficiency scale (OPS). Each of these scales was in a six-point Likert type format. Also, a multiple choice question about the language background of the speaker was included. (The questionnaire is given in Appendix B.)

Raters. The tape and questionnaire were administered to 70 raters (Rs). Half of them were enrolled in undergraduate English composition courses and had neither linguistic nor teaching experience (the naive group), while the other half were

instructors or teaching assistants in ESL (the experienced group). Before taking the test, each R completed a form on biographical data. The naive Rs were tested during their regular class meetings, while the teachers were tested either individually or in small groups.

Rating Procedure. Each sample was rated by the naive native judges and the experienced ESL teachers on each of the six scales. The order of the six scales was the same for all samples of speech. The order of the 16 speakers, however, was randomized for the first tape and was given in reverse order on a second tape. Detailed instructions for the use of the protocols were presented orally with an example using a male Spanish speaker. Form A was used with 15 of the ESL teachers and 20 of the naive judges. Form B was used for 15 of the naive judges and 20 of the ESL teachers. In sum, each of the Rs heard the same directions, either orally or taped, and the same example, and they listened to one of the two tapes of the 16 speakers. They evaluated each speaker on each of the six scales.

RESULTS

Rater Agreement. In order to see whether or not the raters agree in their evaluations of the speaker's accentedness, judges' ratings were treated as variables in a Q-type factor analysis. Normally, of course, the variables input to a factor analysis are test scores, scales, or other measures. In this case, the raters were treated as variables and the various accent scales were treated as subfiles (each containing 16 cases, i.e. the 16 speech samples) in the data set. (Due to computer space limitations, only 60 Rs could be included. Five were excluded from each group on a random basis.) In the first factor analysis, the first four scales were treated together without distinction. We will shortly return to the justification for this (See the discussion of question 4, below.) In the principal components analysis, Factor 1 accounted for 48% of the variance among the individual Rs. All of the Rs loaded positively on this factor and above .36 (Table 1.) Fifty-six Rs showed a correlation of greater than .50 with this factor, while 12 of these 56 loaded at .80 or higher. The overall mean loading was .69. The mean loading for the naive Rs was .645, and for the 30 ESL teachers was .735. On a similar principal components analysis for the OPS, the first factor accounted for 56% of the total variance. The average loading for the naive Rs was .676; for the ESL teachers, it was .816 (Table 2). From these analyses, it can be concluded that there is a very substantial agreement among the Rs, regardless of whether they are naive or experienced.

Difference in Reliability. The second question was whether a significant difference between naive Rs and experienced ESL teachers in the overall reliability of their ratings existed. The answer to this question can be deduced directly from the loadings of the previously defined factors. We can simply contrast the average loading of the naive Rs with the average loading of the ESL teachers. The contrast between the naive and the

Table 1

The First Factor from a Q-Type Principal Components Analysis of the Four Scales Combined (Intelligibility, Pleasantness, Acceptability, and Nativeness).

Naive Rater	Loadings on Factor 1	Experienced Rater	Loadings on Factor 1
6 *	.862	36	.548
7	.500	37	.811
8	.466	38	.691
9	.840	39	.505
10	.849	40	.810
11	.805	41	.778
12	.718	42	.752
13	.729	43	.833
14	.664	44	.666
15	.767	45	.748
16	.580	46	.700
17	.698	47	.664
18	.424	48	.693
19	.721	49	.778
20	.589	50	.476
21	.539	51	.596
22	.791	52	.809
23	.373	53	.739
24	.492	54	.791
25	.694	55	.762
26	.777	56	.804
27	.525	57	.703
28	.444	58	.812
29	.655	59	.741
30	.719	60	.714
31	.687	61	.785
32	.703	62	.681
33	.406	63	.837
34	.547	64	.734
35	.616	65*	.714

*The first and the last five raters were eliminated so as not to exceed the space limitation of the SPSS factor program (PA1, Nie, Hull, Jenkins, Steinbrenner, and Bent, 1975, 479-480).

Table 2

The First Factor from a Q-Type Principal Components Analysis of the Overall Proficiency Scale.

Naive Rater	Loadings on Factor 1	Experienced Rater	Loadings on Factor 1
6*	.952	36	.620
7	.673	37	.731
8	.516	38	.729
9	.797	39	.834
10	.895	40	.863
11	.727	41	.715
12	.848	42	.717
13	.778	43	.914
14	.758	44	.777
15	.830	45	.827
16	.846	46	.804
17	.660	47	.783
18	.361	48	.824
19	.698	49	.850
20	.598	50	.881
21	.642	51	.633
22	.915	52	.873
23	.394	53	.723
24	.437	54	.842
25	.917	55	.792
26	.658	56	.808
27	.615	57	.796
28	.357	58	.841
29	.784	59	.762
30	.570	60	.786
31	.766	61	.801
32	.729	62	.810
33	.326	63	.760
34	.613	64	.828
35	.654	65*	.708

* See the footnote with Table 1.

experienced Rs is not significant at the .05 level for the four scales lumped together, but the contrast for the OPS is signigicant at the .05 level. Therefore, the experienced Rs appear to be somewhat more reliable, although it should be noted that both groups are surprisingly reliable on the whole. This can be inferred from the strenght of the loadings on the two principal factors defined above.

Concurrent Validity. Question 3 concerned the concurrent validity of the accent ratings. A convenient criterion was the speaker's placement at CESL. The overall correlation between the OPS and the various placement levels was .66. For the five Arabic speakers, separately, the correlation was .87. For the five Spanish speakers, it was .71 and for the five Persians, it was only .45. Obviously, the overall ratings were more in agreement for the Arab subgroup.

Dimensions of Accentedness. Question 4 asks whether the several scales are actually sources of unique variance with respect to ratings of accentedness or overall proficiency. Table 2 shows the intercorrelations between the four scales and the OPS with the 70 Rs times the 16 speakers as input data.

Table 3.
Intercorrelations among the Scales of Accentedness

Scales	1	2	3	4	5
1. Intelligibility	1.000	.787	.880	.783	.889
2. Pleasantness		1.000	.786	.727	.790
3. Acceptability			1.000	.792	.892
4. Nativeness				1.000	.802
5. Overall Proficiency					1.000

Each scale appears to be measuring the same thing as the other scale. When mean scores (Table 4) are input for the speakers on each scale the correlation across the scales approximates unity--that is, they are almost perfect, never less than .97.

Table 4.
Mean Correlations of the Scales of Accentedness

Scales	1	2	3	4	5
1. Intelligibility	1.000	.996	.998	.971	.997
2. Pleasantness		1.000	.994	.970	.993
3. Acceptability			1.000	.974	.978
4. Nativeness				1.000	.978
5. Overall Proficiency					1.000

Thus, we may conclude that the scales, in this study, for practical purposes are unitary.

Identification of Language Backgrounds. The fifth question asked how accurately judges would be able to identify each speaker's language background. The experienced Rs correctly identified the source language backgrounds 73% of the time, while the naive Rs could do so only 30% of the time. Identification of the individual language backgrounds by the experienced Rs was best for the Spanish speakers, then the Persians, and then the Arabs. For the naive Rs, the order was Spanish, Arabs, and Persians (Table 5).

Table 5.
The Percentages of Correctly Identified Language Backgrounds

Language	Experienced	Naive
Arabic	66.3	31.4
Spanish	84.6	32.6
Persian	68.6	26.9
Combined	73.0	30.0

Of these three languages only Spanish was studied by more than one R (6 naive Rs and 20 experienced Rs). For the naive Rs it is interesting to note that those who had not had Spanish identified the speakers with 16% greater accuracy than the naive judges who had studied Spanish (Table 6). In looking at the ratings of the experienced Rs, we discover that with or without studying Spanish, Rs could identify the speakers equally well. It seems, therefore, that the ability to correctly identify the language background of a speaker may be due more to mere contact with the speaker of the language in question than to formal study of the language.

Table 6.
The Percentages of Correctly Identified Spanish Speakers

	Experienced	Naive
Those who had studied Spanish	85	20
Those who had not studied Spanish	84	36

DISCUSSION

With little or no research basis, Cartier speculated that teachers are not the most reliable judges of oral proficiency. With equally little empirical study, Jakobovits suggested that naive natives are also not the best judges. Obviously, both Cartier and Jakobovits cannot be right, but neither recommended research to test their claims. With unabashed certainty, language testers

have categorized oral performance into the separate components of accent, grammar, vocabulary, fluency, and comprehension (Valette, 1967; Harris, 1969; Clark, 1972; Heaton, 1975; Davies, 1977). These categories have become the sanctioned criteria for teachers' evaluations of proficiency. Their empirical necessity has gone unquestioned, for the most part.

The data from this experiment disputes both the speculation of Cartier and the opposite claim of Jakobovits; for if they were true, we would expect to find little reliability in the evaluations of either of the two groups. However, the results indicate that both groups can distinguish degrees of proficiency with substantial reliability, although the teachers are somewhat more reliable than the naive judges. On the other hand, the latter fact does not support Jakobovits' claim because of the demonstrated unity of the various scales of accentedness and the overall proficiency rating. Apparently, all of the raters tended to make wholistic unidimensional evaluations, rather than the multi-dimensional and separate evaluations of the presumed components. The unity of the scales suggests that dividing oral performance into components is superfluous at best, and artifactual at worst. According to the available empirical evidence, a listener does not and, indeed cannot componentialize the characteristics of speech. Rather it would appear that the overall comprehensibility is what motivates the evaluation.

The inability of the naive judge to identify the source language backgrounds of the speakers substantiates Palmer's findings (1973). The data, also, show that ESL teachers, the experienced group, are quite sucessful (73%) in identifying source language backgrounds. Further, the data indicate, contrary to Palmer (1973), that accents are quite distinctive.

More research is required before it will be possible to relate degrees of accentedness to points on a well-defined, although subjective, continuum. In addition, experiments should be conducted to see if the ability to identify a speaker's first language affects the reliability and validity of the evaluations. Since studies have shown that speech characteristics may affect personality assessments the converse relationship between personality assessments and speech characteristics on the reliability and validity of evaluations should also be investigated.

A final area of research would be to see how non-natives might evaluate oral performance in ESL. This study used native speakers of English as raters: what if we wanted to generalize to non-natives from different language backgrounds? Would Arabs, for example, tend to rate Arabic speakers in the same way as they might rate native speakers of Spanish or Persian? Would they be more lenient? Would they be as equally reliable?

---

[1] This is an expanded version of a paper presented at The First International Conference on Frontiers in Language Proficiency and Dominance Testing at Southern Illinois University, Carbondale, Illinois or April 22, 1977.

Appendix A

The texts used in this study were selected from several sources. Each passage was rewritten to meet a 100-word criterion. Texts 1, 2, 3, and 4 were each read by two different speakers

1. Ben got off the bus and then the bus drove away. He forgot about the tickets because it was raining. The road was wet and there was a very big hole in his shoe. Then a second bus stopped and he got on. This time there was a seat. He paid a dime for his ticket and then shut his eyes. When he opened them again, the bus was past the theater. He rang the bell and the bus stopped suddenly. It was still raining as he walked back to the theater and went in through the door. He saw many photographs of the actors just before he saw the stage.
(Adapted from Baird, Broughton, Cartwright, and Roberts, 1972, 44.)

2. I hope to learn several foreign languages but English is the one I want to study first. To begin with, I hope to get a good position with one of the big companies in the capital and it will be an advantage for me to have an understanding of English. If my work should ever require my travelling outside of the country- it would be helpful if I knew English. It is used in carrying on business in almost every part of the world. My brothers and sisters, already skillful in English, are eager to practice it with me, so I will have many opportunities when I am ready to speak English.
(Adapted from Van Syoc and Van Syoc, 1971, 89.)

3. On Saturday mornings the big public library opens at half past nine. A lot of the people go into the library on Saturday because this is the time when they go shopping, they take their books into the library, and go home with new ones. Susan and Mary, the two girl librarians, were standing behind the desk. They took the books from the people who came in and gave them their tickets. It was a warm Saturday, and a lot of people were in the streets and in the stores, and many were coming into the library too.
(Adapted from Baird, Broughton, Cartwright, and Roberts, 1972, 47.)

4. As was expected the favorites had gotten well out in front with the remaining horses grouped together some way behind. On a dangerous bend, three of the horses leading the group fell, throwing the riders into great confusion. As the race progressed, the track became full of horses without riders. Towards the end, there were only three horses left. College Joy and Sweet Seventeen were still leading the race with an unknown horse far behind. The crowd was very disappointed when on the last jump in the race, the riders of both favorites failed to keep in the saddle. The crowd cheered and applauded as the unknown horse crossed the finishing line.
(Adapted from Alexander, 1971, 60.)

5. Moving the pilot aside, the man took his seat and listened carefully to the urgent instructions that were being sent by radio from the airport below. The plane was now dangerously close

to the ground, but it soon began to climb. The man had to circle the airport several times in order to become familiar with the controls. The terrible moment came when he had to land the plane. Following the instructions, the man guided the plane towards the airfield. It shook violently as it touched the ground and then moved rapidly across the field, but after a long run it stopped safely.
(Adapted from Alexander, 1974, 61.)

6. The following Sunday we stayed at home, even though it was a fine day. About noon a large and very expensive car stopped outside our house. We were astonished when we saw several people preparing to have a picnic in our small garden. Father got very angry and went out to ask them what they thought they were doing. You can imagine his surprise when he recognized the man who had taken our address the week before. Both men burst out laughing and father welcomed the strangers into the house. In time, we became friends, but we had learned a lesson we have never forgotten.
(Adapted from Alexander, 1974, 63.)

7. It was a very dark and stormy night. Two men were walking slowly down the road. Snow was covering the ground and a cold wind was blowing. They noticed a light behind some trees and soon arrived at a house. A poor old man immediately invited them into a clean room. He seemed a strange fellow, but he spoke kindly and offered them milk and fresh fruit. The men remained there until morning. Then the man led them to the nearest town, but he would not accept any money for his help.
(Adapted from Alexander, 1974, 18.)

8. Science has told us so much about the moon that it is fairly easy to imagine what it would be like to go there. It is certainly not a friendly place. As there is no air or water, there can be no life of any kind. Also for mile after mile, there are only flat plains of dust with mountains around them. Above, the sun and stars shine in a black sky. The moon is very silent. But beyond the horizon, our earth is shining more brightly than the stars. It looks like an immense ball, colored blue, green, and brown.
(Adapted from Alexander, 1974, 35.)

9. The store was empty and very peaceful. We sat down in the main hall and listened to the rain beating against the windows. Suddenly there was a loud noise at the door. Then a large party of boys were led in by a teacher. The poor man was trying to keep them quiet, but they were not paying any attention to him. The boys ran here and there. The teacher explained that the boys were rather excited. But the noise proved too much for us, so we decided to leave. After all, the boys had more right to the store than we did.
(Adapted from Alexander, 1974, 54.)

10. Driving along a highway one dark night, Tom suddenly had a flat tire. Even worse, he discovered that he did not have a spare tire in the back of his car. Tom waved to passing cars and

trucks, but not one of them stopped. At last, he waved to a car like his own. To his surprise, the car actually stopped and a well-dressed woman got out. The woman offered him her spare tire, but Tom had never changed a tire in his life. So she set to work at once and changed the tire in a few minutes while Tom looked on.
(Adapted from Alexander, 1974, 27.)

11. Dan found the school work easy. He read widely both at school and in the branch library. After the third year of high school, he left to take a job with a glass firm. Art work had always been a major interest, and he did so well with the firm that he was promised rapid advancement. But then the depression came, the business failed, and Dan was without a job. At first, he went out looking for a job and continued his art work at home, but when all his efforts brought no results, he stopped looking for work and even lost interest in art.
(Adapted from Whyte, 1955, 8-9.)

12. Tony came into the club to talk the situation over with John. He was trying to get transportation, he said, but even if he could arrange it in the next few minutes it was so late that the boys would miss a large part of the evening. If anyone wanted his money back or a ticket for the next football game, he could have it. John explained the situation to the boys and then said that he thought it would be better if we went another time. Tony agreed. He said that John could collect the tickets later.
(Adapted from Whyte, 1955, 182.)

Appendix B

QUESTIONNAIRE

Name_____ Sex F M
State or Country_____
Native Lg._____ ESL Experience Yrs.__
Other Lgs._____ Mos.__
Age____

In this experiment, you will rate how well some non-native speakers read a short prose passage. In addition, you are asked to identify their native language.

EXAMPLE

NVI (Not Very Intelligible)	1	2	3	4	5	6	VI (Very Intelligible)
UNP (Not Pleasant)	1	2	3	4	5	6	P (Pleasant)
UNA (Not Acceptable)	1	2	3	4	5	6	A (Acceptable)
NN (Non-native)	1	2	3	4	5	6	N (Native)
OPS[4]	1	2	3	4	5	6	
Language Background	Ar.	Sp.	Pr.	Am.	X		

Ar. - Arabic, Sp. - Spanish, Pr. - Persian, Am. - American, X - Unknown

*Overall Oral Proficiency Scale (OPS)

1. The speaker is unintelligible to a native speaker.
2. The speaker has a very heavy accent and makes frequent gross errors.
3. The speaker's accent requires concentrated listening. His mispronunciations lead to occasional misunderstandings.
4. The foreign accent is evident and occasional mispronunciations occur, but these do not interfere with understanding.
5. There are no consistent mispronunciations, but because of occasional deviations would not be taken for a native speaker.
6. The speaker has native pronunciation with no trace of a foreign accent.

Name_____

1. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

2. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

3. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

4. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

5. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

6. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

7. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

8. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

9. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

10. NVI	1	2	3	4	5	6	VI
UNP	1	2	3	4	5	6	P
UNA	1	2	3	4	5	6	A
NN	1	2	3	4	5	6	N
OPS	1	2	3	4	5	6	
LgB	AR	SP	PR	AM	X		

11. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X

12. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X

13. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X

14. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X

15. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X

16. NVI 1 2 3 4 5 6 VI
UNP 1 2 3 4 5 6 P
UNA 1 2 3 4 5 6 A
NN 1 2 3 4 5 6 N
OPS 1 2 3 4 5 6
LgB AR SP PR AM X